Supplemental Material

Bulk Plasmid Carrier Preparation

High-quality, intact, double-stranded, covalently closed, circular form plasmid DNA is essential to avoid undesired sequencing of the carrier. Any free 3' –OH groups in the plasmid might provide binding sites for the polymerase[1] , therefore it is important to eliminate nicked or open plasmids prior to use as a carrier. The two plasmids tested in this study were initially quantified using a Thermo Scientific NanoDrop 8000 spectrophotometer and optical densities were measured to ensure the 260/280 ratio was between 1.8-2.0 and the 260/230 ratio was between 1.7-2.2. 10 micrograms of each plasmid was diluted in 100 microliters of elution buffer. To remove nicked or open plasmids, Exonuclease III, Exonuclease VII, and the reaction buffer from the Exonuclease III kit were added to each plasmid, per kit instructions. The reactions were incubated at 37ºC for one hour, and then returned to 4ºC. The plasmids were purified using two rounds of the 0.6x AMPure bead cleanup procedure, eluted in 100 microliters of elution buffer, and quantified again using the NanoDrop. Using a plasmid carrier similar in size to the target is recommended for optimal processing throughout library construction and sequencing. However, during experimentation using very large circular plasmids (~10kb) as carriers for 20kb libraries, we found that a much smaller proportion survived the Exonuclease digestion. For this reason, the largest plasmid we used as a carrier was pBR322 (4361bp).

Low-Input Shearing

Mechanical shearing of small amounts of DNA to a specific and tight size range is very challenging. This, along with post-shearing cleanup, will continue to be a process that can be further optimized. The purity, quality, and size of the starting DNA all play a role in how the DNA is sheared. Knowing the concentration and average fragment size of a target sample is important for calculating optimal annealing, binding, and loading concentrations. However, when starting with less than 1 nanogram of DNA for shearing, loss of DNA to quantification, in addition to the loss of DNA during shearing and cleanup is not desirable. For this reason, we recommend optimizing shearing conditions prior to attempting to shear an irreplaceable sample. It is important to note that in cases where the starting input amount is less than 1 nanogram, no quality control procedures are used at any point after shearing or during library construction in order to avoid loss of sample.

Shearing to 2kb when using less than 1 nanogram of input DNA requires a reduction in treatment time to the standard protocol used on the Covaris M220. We tested several different time settings using λ genomic DNA as the starting material and found that 600 seconds, instead of the standard 900 seconds, produced the most consistent sizing results. All other settings were standard for the 2kb shear protocol: peak incident power = 8, duty factor = 20%, cycles per burst = 1000, temperature = 7ºC. Recovery rates were ~25-50%, much lower and more variable compared to ~70% recovery when shearing higher amounts of starting material.

Shearing less than 10 ng of DNA to 20kb using the Covaris g-Tube requires a reduction of the centrifuge speed. We found that using 5000 RPM for 60 seconds instead of the standard 6000 RPM produced the most consistent sizing results. Recovery rates were ~30-40% compared to ~70-90% recovery when shearing higher amounts of starting material. Sheared samples were cleaned and concentrated using one round of the 0.6x AMPure bead cleanup procedure. 2kb and 20kb samples were resuspended in 22 microliters and 37 microliters of elution buffer, respectively, for the subsequent end repair or damage repair reactions.

Library Preparation

2kb Low-Input Experiments:

PacBio's standard 2kb Template Preparation protocol was used without any alterations. As a proof of principle, we used λ genomic DNA diluted from a larger bulk shear so that input amounts could be accurately quantified. 500 nanograms of pUC18 plasmid were added to the λ target DNA libraries after ligase inactivation, prior to the addition of Exonucleases III and VII. Following two rounds of 0.6x AMPure bead cleanup, the final libraries were eluted in 10 microliters of Qiagen Elution Buffer.

20kb Low-Input Experiments:

PacBio's standard protocol for Preparing Size-Selected ~20 kb SMRTbell™ Templates was used to prepare 20kb libraries using HB101 *E. coli* DNA that were diluted from a larger bulk shear so that input amounts could be accurately quantified. 2 micrograms of pBR322 plasmid were added to the target sample after ligase inactivation, prior to the addition of Exonucleases III and VII. The lowest cutoff of the High Pass protocol, 4kb, was chosen to maximize the amount of target DNA collected. Following size selection, a final 1.0x AMPure bead cleanup was performed. The final libraries were eluted in 10 microliters of Qiagen Elution Buffer.

Binding Calculator Adjustments and Sequencing

(Go to (SampleNet) for a detailed description)

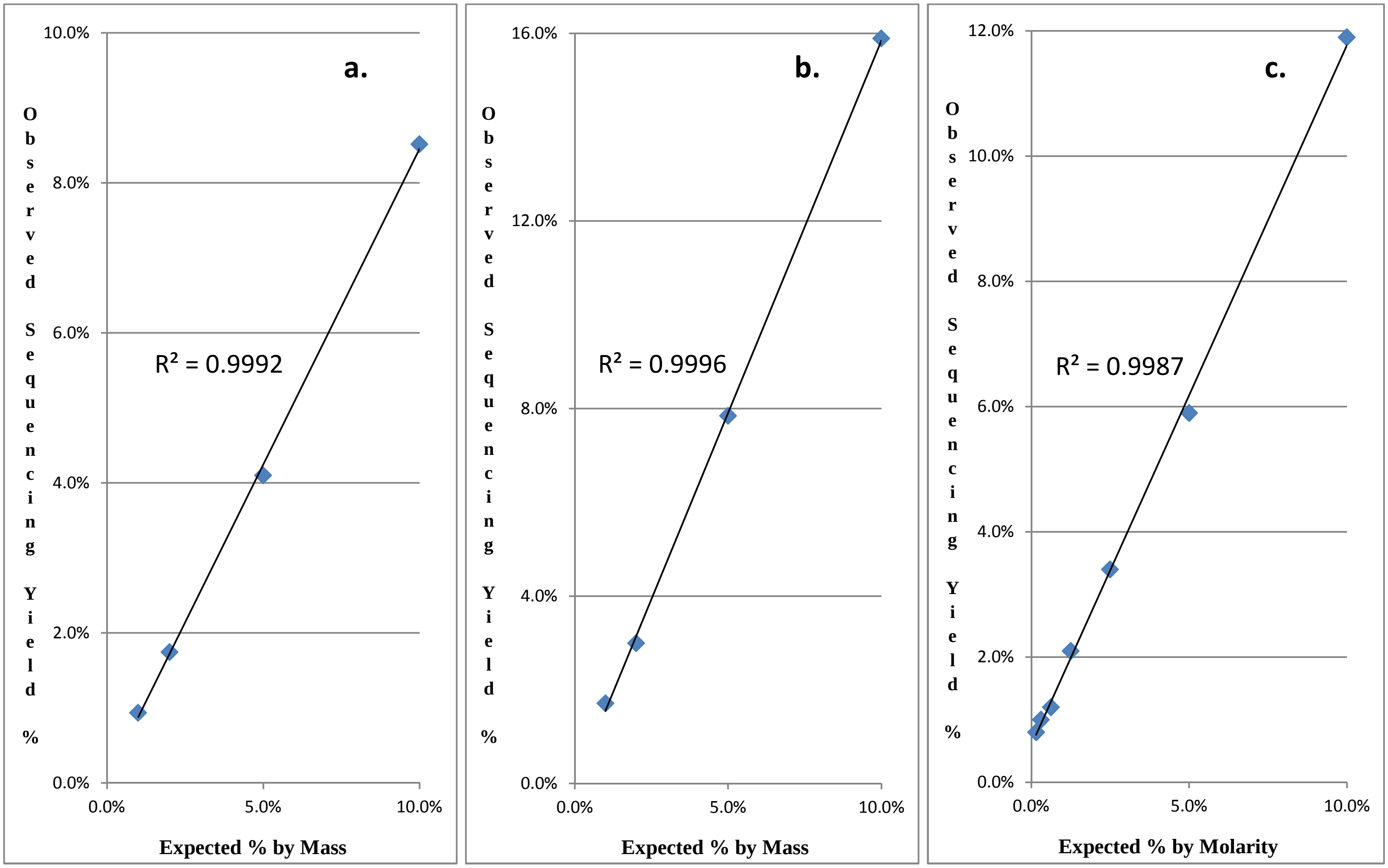
Several adjustments were made to the PacBio annealing and binding calculator to optimize each process and create reagent volumes that could be accurately pipetted by hand. Because the final library quantification step is not performed, thereby avoiding loss of precious material, we enter the average size of the starting material that was used as input in either damage or end repair into the calculator. This is either the expected sheared size based on the selected shearing protocol or if available, the Bioanalyzer-measured size of the starting material. Additionally, we assume 20% recovery of the starting material measured prior to shearing or beginning damage or end repair for 2kb libraries and 10% recovery for 20kb libraries. Based on this assumption, we can calculate the estimated concentration of the final library. Depending on the amount of starting material used, the estimated concentration of the final library is usually so low that the volume of diluted sequencing primer listed in the calculator is too low to be accurately pipetted. To work around this issue, multiply the estimated final library concentration by an offset that results in a diluted sequencing primer volume that is capable of being accurately pipetted. We try to aim for a diluted sequencing primer volume of 1.0 microliter, in an effort to minimize skew to downstream output from the calculator. This can be achieved by dividing 1 by the diluted sequencing primer volume that is currently displayed in the calculator. Multiply the resulting quotient by the estimated library concentration. Enter the resulting product as the new DNA concentration. The diluted sequencing primer volume based on this new DNA concentration should be approximately 1.0 microliter. The total volume of annealed template should go into the polymerase binding reaction. Because the library concentration entered was much higher than what is really present, the calculator says to use only a small portion of bound complex in MagBead binding. However, all bound complex should be used. The optimal salt molarity for MagBead binding is 100mM, but because the majority of salt going into the reaction is 10mM from the bound complex, 100mM Bead Binding Buffer cannot be used. Instead, calculate the volume of 400mM Bead Wash Buffer to add to achieve a 100mM reaction by multiplying the total bound complex volume by 0.3. Add the calculated volume of Bead Wash Buffer to the total volume of bound complex and proceed with the MagBead binding procedure, using a 30 minute incubation time during the 4ºC rotation step. Perform MagBead binding cleanup and sequencing following the standard procedures.

Supplemental Table 1.

Insert Size (bp)	Minimum End-Repair Input (ng)	Average Library Yield (%)	Library Yield (ng)	Library Yield (SMRTbells)*	# of SMRT cells possible ‡	Concentration per SMRT cell (nM)	Mass per SMRT cell (pg)	SMRTbells On Plate	ZMWs per SMRT Cell	SMRTbell: ZMW Ratio
1000	500	20%	100	0.93×10^{11}	992	0.0137	89	0.83×10^{8}	1.5×10^{5}	553
2000	500	20%	100	0.46×10^{11}	496	0.0137	176	0.82×10^{8}	1.5×10^{5}	547
5000	2000	20%	400	0.74×10^{11}	645	0.0169	548	1.02×10^{8}	1.5×10^{5}	680
10000	5000	15%	750	0.70×10^{11}	605	0.0169	1093	1.01×10^{8}	1.5×10^{5}	673

* Number of molecules = (ng * number/mole) / (bp * ng/g * g/ mole of bp) ‡MagBead Loading P4/C2 library input requirements, expected average yield, concentration on plate, and theoretical SMRTbell to ZMW ratio.

Supplemental Figure 1.



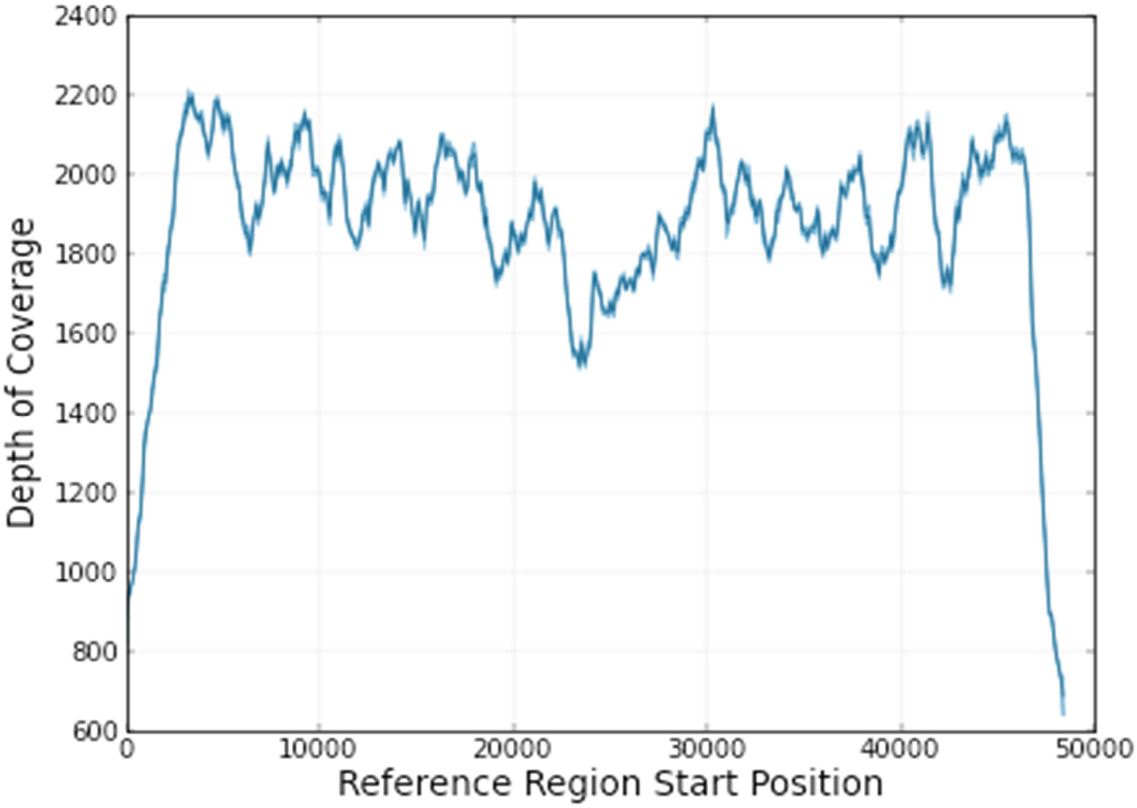
a. Observed Percentage of Low-Input Target Sequencing Yield vs. Expected Percentage by Mass: 500bp libraries
b. Observed Percentage of Low-Input Target Sequencing Yield vs. Expected Percentage by Mass: 2kb libraries
c. Observed Percentage of Low-Input Target Sequencing Yield vs. Expected Percentage by Molarity: 2kb libraries

Supplemental Table 2.

End-Repair Input (ng)	Amount Recovered Assuming 20% Recovery (pg)	Required Mass On Plate (pg)	SMRTbells On Plate	ZMWs per SMRT Cell	SMRTbell: ZMW Ratio
50	10,000	176	0.82×10^{8}	1.5×10^{5}	547
5	1,000	176	0.82×10^{8}	1.5×10^{5}	547
1	200	176	0.82×10^{8}	1.5×10^{5}	547
0.5	100	176	0.41×10^{8}	1.5×10^{5}	274
0.05	10	176	0.41×10^{7}	1.5×10^{5}	27
0.005	1	176	0.41×10^{6}	1.5×10^{5}	2.7

2kb λ Low-Input Experiment: Decreasing input amounts, expected theoretical recoveries by mass and number of molecules, and theoretical SMRTbell to ZMW ratio.

Supplemental Figure 2.



Observed depth of coverage across lambda_NEB3011 (window size = 100bp).
2kb λ Input Amount = 500 picograms

Supplemental Table 3.

Input Amount (ng)	Post-Filter Yield (Mb)	Target Mapped Yield (Mb)	Carrier Mapped Yield (Mb)	Depth of Coverage from Subreads	Target Size (bp)	CCS Reads
0.47*	158.0	89.97	0.044	63,272	1422	12,219
0.50	342.4	197.05	0.004	152,633	1291	12,655
0.50	150.0	94.22	0.143	67,063	1405	4,719
2.70	136.0	51.22	0.016	35,707	1434	1,863
2.70	232.5	124.54	0.019	86,795	1434	5,371
2.70	171.6	99.99	0.014	70,290	1422	2,859
5.00*	397.4	217.04	0.006	155,139	1399	24,247
5.00*	440.6	271.7	0.007	195,611	1389	31,684

*P4/C2-120 minute movie

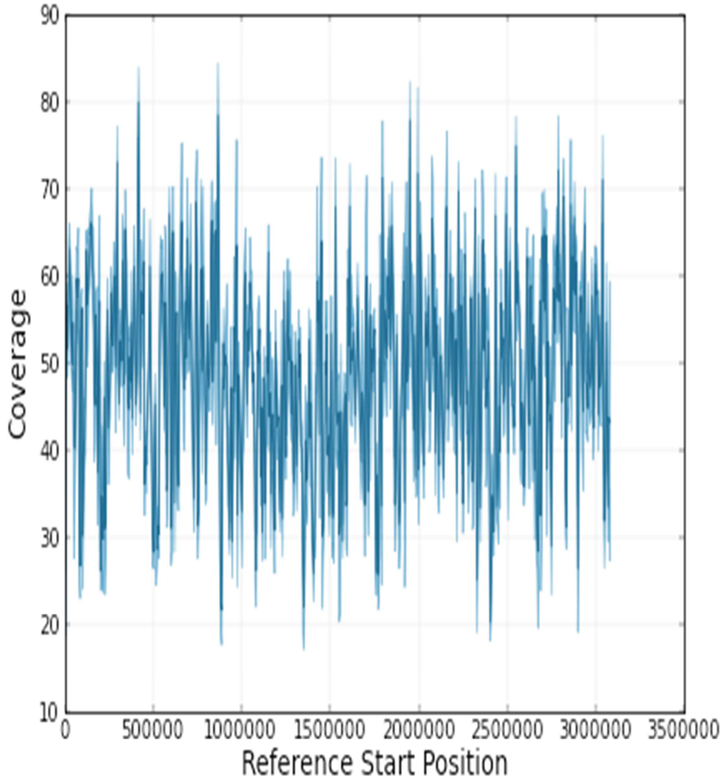
Per SMRTcell sequencing statistics produced from “real world” samples using the 2kb low-input protocol.

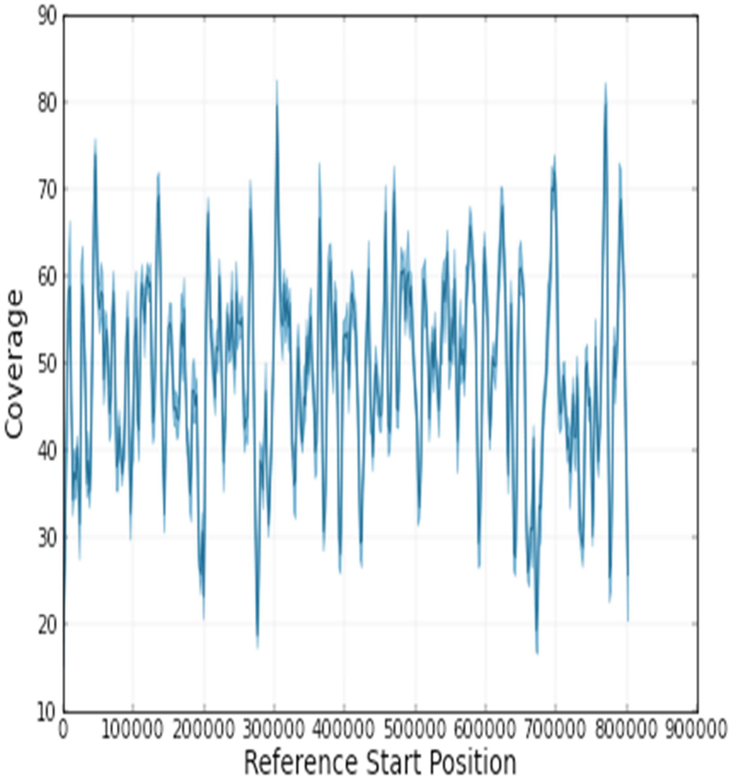
Supplemental Table 4.

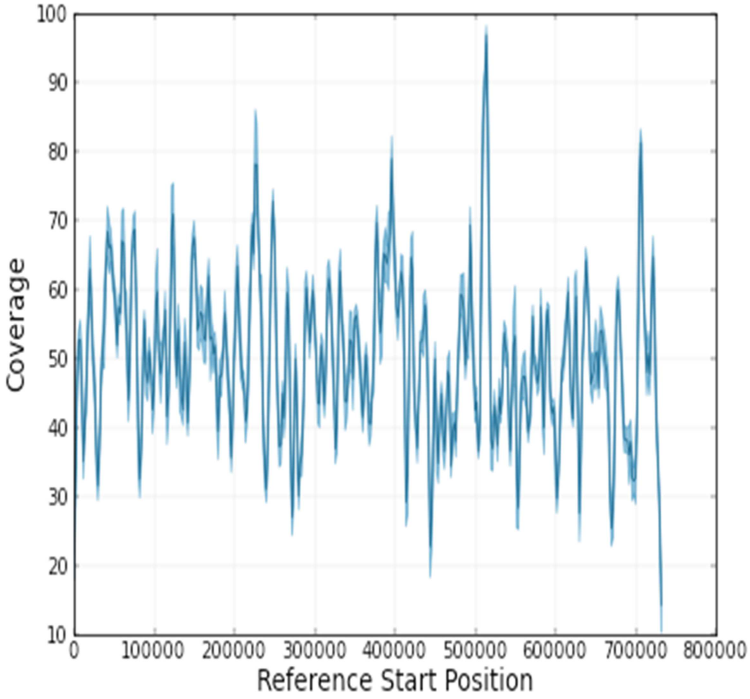
Input Amount (ng)	Number of Uses	Post-Filter Yield (Mb)	Target Mapped Yield (Mb)	Carrier Mapped Yield (Mb)	Mapped Rd_Ln Of Insert (bp)	Mapped Rd_Ln 95% (bp)	Depth of Coverage from Subreads	Polished Contigs	N50 Contig Length	Sum of Contig Lengths	Consensus Con-cordance
50	2	306.8	203.4	0.010	3,807	10,103	36.2	34	233,984	4,638,363	99.985
20	3	289.9	244.4	0.014	4,316	10,525	48.6	5	3,088,137	4,640,431	99.986
10	3	171.4	142.3	0.061	3,900	9,484	29.3	139	61,389	4,420,920	99.960
5	3	149.1	127.0	0.056	4,220	11,146	25.5	113	82,071	4,581,970	99.957

HGAP2 results summary from 20kb low-input experiments using decreasing amounts of *E.coli* target in plasmid carrier.

Supplemental Figure 3.







Observed depth of coverage across three large contigs produced from 20 nanograms of *E. coli* DNA using the 20kb low-input protocol.